No Evidence for Bilingual Cognitive Advantages: A Test of Four Hypotheses

Claudia C. von Bastian

University of Zurich

University of Colorado Boulder

Alessandra S. Souza

Miriam Gade

University of Zurich

Author Note

Claudia C. von Bastian, Department of Psychology and Neuroscience, University of Colorado Boulder, United States; Alessandra S. Souza, Department of Psychology, University of Zurich, Switzerland; Miriam Gade, Katholische Universitaet Eichstaett-Ingolstadt, Germany.

This work was supported by a grant from the Suzanne and Hans Biaesch Foundation for Applied Psychology. We thank Myriam Leupin, Sarah Schoch, Michael Uehlinger, and Alisa Schaettin for their assistance with collecting the data. We also thank Akira Miyake, Nicholas Carruth, Daniel Gustavson, and John Lurquin for their comments on a draft of this manuscript.

Correspondence concerning this article should be addressed to Claudia C. von Bastian, Department of Psychology and Neuroscience, UCB 345, Boulder, CO 80309, United States. Fax: +1 303 492 1234. E-mail: claudia.vonbastian@colorado.edu

Abstract

The question whether being bilingual yields cognitive benefits is highly controversial with prior studies providing inconsistent results. Failures to replicate the bilingual advantage have been attributed to methodological factors such as comparing dichotomous groups and measuring cognitive abilities separately with single tasks. Therefore, we evaluated the four most prominent hypotheses of bilingual advantages for inhibitory control, conflict monitoring, shifting, and general cognitive performance by assessing bilingualism on three continuous dimensions (age of acquisition, proficiency, and usage) in a sample of 118 young adults, and relating it to nine cognitive abilities each measured by multiple tasks. Linear mixed-effects models accounting for multiple sources of variance simultaneously and controlling for parents' education as an index of socio-economic status revealed no evidence for any of the four hypotheses. Hence, our results suggest that bilingual benefits are not as broad and as robust as has been previously claimed. Instead, earlier effects were possibly due to task-specific effects in selective and often small samples.

Keywords: bilingual advantage, executive functions, cognitive control

No Evidence for Bilingual Cognitive Advantages: A Test of Four Hypotheses

A multitude of studies suggests that being bilingual does not only enhance language control, but also yields non-linguistic cognitive performance benefits (for a review, see Bialystok, Craik, & Luk, 2012). These benefits are assumed to result from the life-long practice in dealing with multiple simultaneously active languages (for an overview, see Kroll, Bobb, & Hoshino, 2014), and are expected to occur particularly for executive functions (EF, i.e., cognitive processes regulating thoughts and behavior, Miyake & Friedman, 2012). Indeed, bilingual advantages have been documented for all three abilities typically subsumed as EF: inhibiting prepotent responses (e.g., Bialystok, Craik, Klein, & Viswanathan, 2004; Martin-Rhee & Bialystok, 2008), shifting between mental task sets (e.g., Prior & MacWhinney, 2010; Wiseheart, Viswanathan, & Bialystok, in press), and updating of working memory contents (e.g., Bialystok, Craik, & Luk, 2008; Luo, Craik, Moreno, & Bialystok, 2013).

However, recent failures to replicate earlier findings (e.g., Gathercole et al., 2014; Kirk, Fiala, Scott-Brown, & Kempe, 2014; Kousaie, Sheppard, Lemieux, Monetta, & Taler, 2014; Morton & Harper, 2007; Paap & Greenberg, 2013; Paap, Johnson, & Sawi, 2014) in combination with a publication bias favoring the report of significant effects (de Bruin, Treccani, & Della Sala, 2015) have led to an intense debate about the very existence of the effect (e.g., see Paap, 2014; Valian, 2015). Multiple methodological factors have been discussed as potentially causing these divergent findings. In this study, we therefore tested the four most prominent hypotheses of bilingual advantages while accounting for the methodological issues in past research.

Four Hypotheses of Bilingual Advantages

Each of the four hypotheses predict bilingual advantages for different (although sometimes overlapping) cognitive abilities, depending on the assumed underlying mechanism (see Table 1). The *Inhibitory Control Advantage* (referred to as "BICA" by Hilchey & Klein,

2011) is assumed to result from the bilinguals’ constant practice in exerting control processes to inhibit the language currently not in use (cf. the Inhibitory Control Model, Green, 1998). More specifically, this hypothesis predicts that bilinguals show smaller interference effects in inhibition tasks than monolinguals.

Following observations that bilinguals do not necessarily differ from monolinguals in the magnitude of the interference effect, but instead are overall faster in tasks requiring conflict resolution (e.g., Costa, Hernández, & Sebastián-Gallés, 2008; Martin-Rhee & Bialystok, 2008), it has been hypothesized that bilinguals may benefit from generally enhanced conflict monitoring processes (also referred to being a bilingual executive processing advantage, "BEPA", Hilchey & Klein, 2011). These benefits are assumed to be caused by the bilinguals’ constant need to select the appropriate language depending on the current interlocutor, thereby forcing them to continuously monitor their environment for conflicting information (e.g., Costa, Hernández, Costa-Faidella, & Sebastián-Gallés, 2009). More specifically, the *Conflict Monitoring Advantage Hypothesis* predicts that a higher degree of bilingualism yields enhanced performance in tasks entailing conflict, even in those task conditions without conflict (i.e., neutral or congruent trials, cf. Hilchey, Saint-Aubin, & Klein, in press).

Based on their practice in switching between languages, the *Shifting Advantage Hypothesis* (cf. Bialystok, et al., 2012) predicts that bilinguals show lower costs in reaction times when switching between two tasks than monolinguals. Finally, the *Generalized Cognitive Advantage Hypothesis* assumes that the bilingual experience is so profound that it alters cognitive processing in general, resulting in overall higher mental flexibility (Kroll & Bialystok, 2013). It therefore predicts that effects of bilingualism may be difficult to detect in single abilities, but are more likely to emerge in the shared variance across multiple fluid cognitive abilities (e.g., executive functions and reasoning).

Potential Sources of Inconsistencies between Findings

Failures to observe bilingual advantages have been criticized on several grounds. First, participants are often assigned to dichotomous groups (i.e., monolinguals or bilinguals). However, it has been suggested that bilingualism is a multidimensional, continuous phenomenon and, hence, should be treated as such (Kroll & Bialystok, 2013; Luk & Bialystok, 2013). Second, past studies often used only single indicators measuring a single EF. Using single indicators poses the problem of task-impurity (Miyake & Friedman, 2012) and makes it difficult to distinguish between task-specific and ability-general effects (cf. Shipstead, Redick, & Engle, 2012). Moreover, it has been argued that the bilingual experience affects cognitive performance so profoundly yet generally that effects could go undetected when investigating only one EF at a time (Kroll & Bialystok, 2013).

Third, many previous studies lack control of potential confounds that might obscure or mimic effects of bilingualism such as socioeconomic status (SES, cf. Fuller-Thomson & Kuh, 2014; Kousaie & Phillips, 2012; Morton & Harper, 2007) or leisure activities that are also potentially beneficial for cognitive performance such as video-gaming, musical training, or doing exercise (Valian, 2015). Last but not least, evidence favoring the bilingual advantage stems primarily from small-scale studies (average n = 29), whereas larger-scale studies (average n = 45; Paap, et al., 2014) more often reported null-effects (e.g., Antón et al., 2014; Duñabeitia et al., 2014; Gathercole, et al., 2014; Hernández, Martin, Barceló, & Costa, 2013). This is particularly critical as small samples have been shown to be more prone to produce false-positive findings (Button et al., 2013).

Current Study

In the present study, we tested the four hypotheses of bilingual advantages with at the same time aiming at addressing the methodological issues outlined above. To test the hypotheses outlined in Table 1, we assessed nine cognitive abilities: Inhibition, Conflict Speed, Mixing, Working Memory (WM) Monitoring, Shifting, Updating, WM Capacity, and Reasoning.

The *Inhibitory Control Advantage Hypothesis* was evaluated with well-established inhibition tasks (e.g., see Hilchey & Klein, 2011), from which we derived the cost in reaction time (RT) for suppressing predominant responses. The *Conflict Monitoring Advantage Hypothesis* was primarily examined regarding speed in the context of conflict (hereafter referred to as Conflict Speed), which was indicated by mean RTs in the baseline conditions (i.e., congruent or neutral trials) in the inhibition tasks. Furthermore, we evaluated whether potential benefits in conflict monitoring would extend to Mixing and WM Monitoring. Mixing reflects the monitoring of multiple task goals and was assessed with the task switching paradigm (Monsell, 2003) by comparing performance in single-task blocks and mixed-task blocks. Reduced time costs for mixing between task-sets has also been observed in bilingual samples compared to monolinguals (cf. Soveri, Rodriguez-Fornells, & Laine, 2011; Wiseheart, et al., in press). WM Monitoring refers to coordinating single information elements into novel structures in WM, and to detect whenever these structures form a critical constellation while suppressing irrelevant context information (e.g., Oberauer, Süß, Schulze, Wilhelm, & Wittmann, 2000). The *Shifting Advantage Hypothesis* was tested on switch costs (hereafter referred to as Shifting) derived from the task switching paradigm. As bilingual benefits have also been discussed for Updating and WM Capacity (e.g., Engel de Abreu, 2011; Luo, et al., 2013), and Reasoning (e.g., Hakuta & Diaz, 1985), our assessment also included tasks measuring these abilities. Finally, the *Generalized Cognitive Performance Advantage Hypothesis* was evaluated inspecting the main effects of bilingualism across all cognitive abilities assessed. Processing Speed served as a baseline measure.

We addressed the methodological issues in past research in three ways. First, we examined three widely used indicators of bilingualism as continuous predictors of cognitive abilities: age of acquisition, usage, and proficiency. In addition to their unique contributions, we examined interactions between these three indicators by forming data-driven clusters of participants with different levels of bilingualism using a cluster-analytic approach. Second, we evaluated the effects of bilingualism on the nine cognitive abilities assessed simultaneously, and measured each of them with multiple tasks. Third, we included measures of parents' education (as a proxy for SES) and leisure activities as control variables. Fourth, we tested the bilingual advantage with a comparatively large sample (N = 118).

Method

Participants first completed a computer-based test battery measuring nine cognitive abilities, followed by questionnaires assessing language, demographic, and socioeconomic background (i.e., parents' education). Participants were tested in our laboratory in groups of up to five in one 4.5 h session (including three 10 min breaks). To control for linear effects of fatigue and practice, half of the participants completed the test battery in reversed order than the other half.

Participants

We aimed at a sample size of N = 120, which we determined based on studies by members in our laboratory that investigated individual differences (e.g., Oberauer, et al., 2000; Oberauer, Süß, Wilhelm, & Wittmann, 2003; von Bastian & Oberauer, 2013). Students from Swiss universities were recruited for a study examining the impact of language experience on cognition. The experimental protocol was approved by the institutional review board at the University of Zurich. All participants gave written consent to taking part in the study. We collected data from a total of N = 121 participants, because one participant's data had to be replaced after an experimenter error. Two additional participants were excluded from the final data set because they did not provide the age of acquisition of their languages. Demographics of the remaining 118 participants who were included in the statistical analyses are listed in Table 2. All participants reported normal or corrected to normal vision, and none of them exhibited evidence of colorblindness as determined by Ishihara’s color test (Ishihara, 2003). At study completion, participants received CHF 40 (about USD 40) or course credits.

Switzerland is officially a multilingual country, but Zurich is located in midst of the German speaking part of Switzerland. Hence, despite the obligatory language learning at school – the education system requires that two languages (English and one of the other official Swiss languages such as French) are taught – most people use German as their language at home as well as at work (Bundesamt für Statistik, 2015b). Language use is therefore comparable to monolingual regions in other multilingual countries such as Canada or Spain. About a third of the study sample reported to have immigrated to Switzerland (see Table 2), which is representative for the overall proportion of immigrants in the Swiss population (Bundesamt für Statistik, 2015a). The majority of participants with immigration status in the language cluster with the lowest degree of bilingualism indicated to possess German or Austrian citizenship, two neighboring countries in which the same language is spoken as in Zurich (i.e., German).

Most participants (n = 91) reported German as their first language (L1). For the remaining 27 participants, L1 was diverse: French (3), Italian (3), Turkish (3), Polish (3), Serbo-Croatian (3), English (2), Spanish (2), Hungarian (2), Bulgarian (1), Chinese (1), Hindi (1), Malayalam (1), Vietnamese (1), and Portuguese (1). Likewise, a variety of second languages (L2) was reported: German (23); French (44); English (33); Italian (4); Chinese (2), Dutch (2), Greek (2), Spanish (2), Hungarian (1), Romansh (1), Russian (1), Slovenian (1), Thai (1), and Urdu (1).

Cognitive Assessment

Each of the nine cognitive abilities assessed was operationalized by at least three dependent variables covering figural-spatial, verbal, and numerical material, resulting in a total of 21 computer-based tasks[1]. Each of the tasks started with several practice trials preceding test blocks of pseudo-randomized trials (see Table 3 for the number of trials for each task). All dependent measures were coded so that positive values reflect better performance. Tasks were implemented with Tatool, and open-source framework for programming psychological experiments (von Bastian, Locher, & Ruflin, 2013).The executive functions tasks reported here are available on Tatool Web (www.tatool-web.com).

Inhibition and Conflict Speed. This set of tasks measures the suppression of predominant responses. In the *Simon* task, participants were presented a circle on either side of the screen. Each trial started with a fixation cross centrally presented for 250 ms before the circle was displayed. Participants were then asked to press the left arrow key when the circle was green and the right arrow key when the circle was red. Hence, the spatial location in which the circle appeared and the spatial location of the response could be either congruent (e.g., a green circle appearing on the left, 75% of the trials) or incongruent (e.g., a green circle appearing on the right, 25% of the trials). As demands of inhibitory control have been shown to be higher with a decreasing proportion of incongruent trials (Logan & Zbrodoff, 1979), only 25% of the trials in this task were incongruent.

In the *Flanker* task, participants had to decide whether a centrally presented target was a vowel (A or E) or consonant (S or T). Stimuli were presented until the participants' indicated their response by pressing the left (i.e., vowel) or right (i.e., consonant) arrow key. The target was flanked by three irrelevant stimuli on each side. The flankers were either congruent (i.e., from the same category as the target, for example AAAEAAA), incongruent (i.e., from the other category than the target, for example AAATAAA), or neutral (i.e., unrelated to the target's category, for example ###S###). Congruent, incongruent, and neutral trials occurred in an equal proportion of the trials.

In the *Stroop* task, participants had to count the number of 1 to 4 centrally displayed characters and indicate their response by key press. In congruent trials, the number of characters corresponded to the digits displayed (e.g., 333), whereas in incongruent trials, the magnitude of the digits displayed differed from the number of characters (e.g., 44). Neutral trials were unrelated symbols (e.g., ###). All three trial types occurred with equal frequency.

To measure Inhibition, we computed proportional RT interference scores by subtracting RTs in the baseline condition (congruent trials in the Simon task, and neutral trials in the other two tasks[2]) from RTs in incongruent trials and dividing the result by the baseline RTs. Mean RTs in the baseline condition served as a measure for Conflict Speed.

Mixing and Shifting. The task switching paradigm (Monsell, 2003) was used to assess Mixing and Shifting. In these tasks, participants have to classify stimuli according to specific rules. In the *Color/Shape* task, participants classified simple geometrical shapes regarding their color (green or blue) or shape (round or angular). The 64 Stimuli were composed of 16 round and 16 angular shapes, with each shape being used once in green and once in blue color. In the *Animacy/Size* task, participants classified line-drawings of simple objects or animals regarding their animacy (living or non-living) or size (smaller or larger than a soccer ball). Line-drawings were 32 animals and 32 simple objects from Snodgrass and Vanderwart (1980) and Szekely et al. (2004). Half of the animals and objects were smaller (e.g., a worm or a comb), and the other half of it being larger than a soccer ball (e.g., a giraffe or a boat). In the *Parity/Magnitude* task, participants classified digits (1-9 excluding 5) regarding their parity (even or odd) or magnitude (smaller or greater than five).

Each task consisted of five blocks: two single-rule blocks (e.g., color classification followed by shape classification) preceding a mixed-rules block (e.g., switching between color and shape classification), which was then followed by the two single-rule blocks in reversed order (e.g., shape classification followed by color classification). A visual cue (see Figure 1) presented 150 ms before stimulus onset indicated the rule that had to be applied. In single-rule blocks, the same rule had to be applied throughout the block. In the mixed-rules block, the two rules switched unpredictably with half of the trials being repetition trials (two successive trials in which the same rule had to be applied) and the other half being switch trials (the rule changed from the preceding to the current trial).

To assess Mixing, RTs in single-rule blocks were subtracted from RTs in repetition trials in the mixed-rules block, the result of which was divided by their average. Shifting was indicated by switching costs, which were derived from the mixed-rules block by subtracting repetition RTs from switch RTs and dividing the result by their average RT. Average RTs in the single-rules block served as a measure of Processing Speed.

WM Monitoring . We used the monitoring tasks from von Bastian and Oberauer (2013, based on Oberauer, et al., 2003). Participants had to watch several independently changing objects and to respond when a certain critical relation between the objects occurred. In the *Squares* task, 2 of 20 dots in a 10 x 10 grid randomly changed their position within the grid every 2 s. Participants had to press space whenever four dots formed a square. In the *Rhymes* task, 1 of 9 words in a 3 x 3 grid changed every 2 s. Here, participants had to press space whenever three words in a row (horizontal, vertical, or diagonal) rhymed. In the *Digits* variant, 1 of 9 three-digit numbers in a 3 x 3 grid changed every 2 s. Participants had to press space whenever three numbers in a row had the same last digit. Each task consisted of 16 runs, each of which comprised 2–8 changes until the critical relation occurred. The dependent measure was detection performance, which is $d' = z(\mathrm{H}) - z(\mathrm{FA})$, where *H* is the hit rate, *FA* the false alarm rate, and *z* refers to the *z*-value corresponding to the probability of the given argument.

Updating. Tasks used to measure Updating were modeled after the keep-track tasks developed by Miyake et al. (2000). Participants had to memorize an initial set of stimuli and update these memoranda in case new information was presented. After nine updating steps, participants were asked to recall the most recent value of each memorandum. To ensure that the initial set of memoranda had to be encoded, in 5 out of 25 trials recall was probed immediately after the initial encoding. In the *Color Keep-Track* task, the colors of five shapes (circle, square, triangle, diamond, and hexagon) had to be updated. After the updating phase, participants had to recall the color each shape has been last presented in by selecting the correct one out of the ten possible options. In the *Letter Keep-Track* task, the specific letter displayed in 1 of 5 boxes had to be updated. In the Number Keep-Track task, four digits were presented in different colors (red, blue, green, and orange). During the updating phase, the value of each digit could change (ranging from 1 to 9). Accuracy (i.e., number of correctly solved items divided by the total number of items) was used as dependent measure.

WM Capacity. To measure figural-spatial working memory capacity, we used the *Spatial Short-Term Memory* task described in Lewandowsky, Oberauer, Yang, and Ecker (2010). Participants had to remember the relation between the spatial locations of a series of sequentially presented dots in a 10 x 10 matrix. Participants had to reproduce the relative position of the dots in the matrix. The number of dots in each pattern ranged from 2 to 6. In the *Brown-Peterson* task, participants first had to memorize 3 to 6 sequentially presented words. Next, they had to complete five trials of a distractor task, in which lexical decisions on four-character strings had to be made. Finally, at the end of the each trial, all memoranda had to be recalled in their correct serial order. In the *Complex Span* task, participants had to memorize 3 to 6 two-digit numbers. Presentation of the memoranda was interleaved by a distractor task in which participants were asked to judge the veracity of equations (e.g., “1 + 3 = 5”). At the end of each trial, memoranda had to be recalled in correct serial order.

In case of the Spatial Short-Term Memory task, the dependent measure was the proportion of correctly recalled relations in the pattern. For this purpose, the patterns reconstructed by the participants were matched against the original pattern. In each trial, participants could receive 1 point for any dot that deviated only one cell in any direction from the original, and 2 points for a perfect match (for more details on this scoring procedure, see Lewandowsky, et al., 2010). For the Brown-Peterson and the Complex Span task, the proportion of items recalled at the correct position (partial-credit unit score, cf. Conway et al., 2005) served as dependent measure.

Reasoning. We administered five time-restricted reasoning tests. In Arthur and Day’s short version (1994; see also Arthur, Tubre, Paul, & Sanchez-Ku, 1999) of *Raven's Advanced Progressive Matrices* (Raven, 1990), participants had to complete a pattern by choosing one of eight alternatives. In the *Locations Test* (Ekstrom, French, Harman, & Dermen, 1976), participants had to discover the rule of patterns of dashes, one of which in each of four rows was replaced by an “x”. Participants had to select the correct location of the next “x” out of five alternatives. In the *Letter Sets Test* (Ekstrom, et al., 1976), five sets of four letters were presented that all followed a certain logical pattern except for one set. Participants had to select the deviating letter set. In the *Nonsense Syllogisms Test* (Ekstrom, et al., 1976), the task was to decide whether conclusions drawn from two premises with nonsensical content were logically valid (e.g., following the premises “all trees are fish” and “all fish are horses”, it would be logically correct to conclude that “therefore all trees are horses”). In the *Diagramming Relationships* (Ekstrom, et al., 1976) task, participants had to choose which out of five diagrams represents best a set of three given nouns. For example, the set “animals, cats, and dogs” would be best represented by one circle corresponding to “animals” containing two separate circles for “cats” and “dogs”. The dependent measure for all Reasoning tasks was the number of correctly solved items divided by the total number of items.

Language and Background Assessment

After completing the test-battery, participants filled a custom-made questionnaire assessing their language background, leisure activities, demographics, and SES. In addition, participants completed the German version of the Language Experience and Proficiency Questionnaire (LEAP-Q, Marian, Blumenfeld, & Kaushanskaya, 2007) from which we derived age of L2 acquisition, proportion of daily non-L1 language usage, and L2/L1 proficiency ratio.

Data Analysis

Data Preprocessing

Only RTs of correct responses were analyzed. RTs of responses following immediately after wrong responses, and RT outliers were excluded from analyses. Outliers were defined as RTs being 3 median absolute deviations away from the overall median (Leys, Ley, Klein, Bernard, & Licata, 2013). All dependent variables were *z*-transformed, and continuous predictors were grand-mean centered. Task descriptive statistics and reliabilities are listed in Table 3, and zero-order between-tasks correlations are listed in Table S1 in the Supplemental Material.

Randomly allocating participants to two different orders of test administration is useful to control for fatigue and practice effects in long testing sessions such as the present one. However, it introduces an unwanted source of variance. Therefore, we eliminated this variance by arbitrarily choosing one order as the reference condition and correcting the data of the other order for the mean difference between the two orders for each variable (cf. von Bastian & Oberauer, 2013).

Cluster Analysis

To investigate whether particular combinations of the three indicators of bilingualism (i.e., possible interactions thereof) can explain individual differences in cognitive performance while still keeping the complexity on an interpretable level, we ran a k-means cluster analysis reducing the three bilingual variables to a single indicator. We used the “kmeans” function implemented in R (R Core Team, 2014). The k-means procedure creates *k* groups by minimizing the within-group variance and maximizing the between-group variance. The k-means algorithm is based on the Euclidian distance and, thus, it is sensitive to the scaling of the variables used in the clustering procedure. Therefore, we used the z-transformed indicators even though the non-transformed scores converged perfectly (i.e., participants were classified into exactly the same clusters). The k-means procedure requires pre-specifying the number of clusters, which we determined by plotting the total within-groups sums of squares against the number of clusters (which we allowed to vary between 1 and 15). The bend in the scree plot is considered informative in defining how many cluster solutions should be imposed on the data. It favored a three-cluster solution (see Figure 1). We interpreted these three language clusters (see also Figure 1) as groups of participants with low, medium, and high level of bilingualism.

We validated the language clusters based on other language variables that were assessed but not used for specifying the clusters. Language clusters differed in terms of the locus of L2 acquisition (formally at school vs. in family context), $\chi^2(2, n = 118) = 24.77$, $p < .001$, and whether they considered themselves as simultaneous bilinguals (i.e., indicated L2 as a second native language) or sequential bilinguals, $\chi^2(4, n = 118) = 26.74$, $p < .001$. In contrast to the language variables, it is desirable that the clusters should not differ in demographics. This was precisely the case, as language clusters were comparable in terms of gender ($\chi^2(2, n = 118) = 0.17$, $p = .916$), age (Welch $F(2, 65.3) = 0.82$, $p = .444$), years of education (Welch $F(2,64) =$

1.76, $p = .180$), and parents' education ($F(2, 69.3) = 1.44$, $p = .243$). In addition, we evaluated whether language clusters differ systematically in leisure variables recently discussed as potential confounds (Valian, 2015). We observed no differences for self-reported hours per week spent on musical training (Welch $F(2, 57.5) = 1.41$, $p = .254$), video gaming (Welch $F(2, 68.9) = 0.80$, $p = .452$), or physical exercise (Welch $F(2, 65.3) = 2.23$, $p = .116$).

Linear-Mixed Effects Modeling

We ran linear mixed-effects (LME) models to evaluate the impact of bilingualism on the level of cognitive ability rather than for single tasks (cf. von Bastian & Oberauer, 2013). LME models account for multiple sources of variance simultaneously, which can be specified as fixed effects (which account for the effect of experimental conditions or predictors) or random effects (which account for the variability in sampling of, for example, individuals or tasks). We ran two types of models: Model 1 containing continuous predictors of bilingualism (age of acquisition, non-L1 language usage, and proficiency ratio), and Model 2 with cluster membership as categorical predictor. In addition, both models included the two categorical predictors cognitive ability (each of eight abilities contrasted to Processing Speed as baseline), and task material (figural-spatial vs. verbal-numerical task), as well as their interactions with the indicators of bilingualism. Both models also included parents' education as a covariate, but results were essentially the same when excluding this predictor (see Tables S2 and S3 in the Supplemental Material).

We included two crossed random effects (Baayen, Davidson, & Bates, 2008) for subject and task in the models, reflecting that both the participants and the tasks included in our study are only samples drawn from larger populations. Based on a recent simulation study (Barr, Levy, Scheepers, & Tily, 2013), we chose the most anti-conservative random-effects structure allowing random variation of subject and task on the intercepts only. Continuous predictors were centered, and categorical predictors were dummy-coded with the intercept representing the baseline mean. Models were fit in R (R Core Team, 2014) with the package “lme4” (Bates, Maechler, Bolker, & Walker, 2014). Satterthwaite approximation implemented in the package “lmerTest” (Kuznetsova, Bruun Brockhoff, & Haubo Bojesen Christensen, 2014) served to estimate the degrees of freedom. The significance of predictors was determined using an alpha-level of .05.

Results

First, we examined the effects of bilingualism on single-task performance. Correlations between the continuous indicators of bilingualism and single-task performance are listed in Table 4. Note that for age of acquisition, higher values reflect an older age when L2 was acquired, and thus negative correlations indicate better cognitive performance with a higher degree of bilingualism, while the opposite is the case for usage and proficiency.

Effects of language cluster on single tasks alongside means and standard deviations for the three language clusters are presented in Table 5. Correlations and t-tests between language clusters indicated only selective effects of bilingualism on single tasks (positively for the Simon Inhibition and color-shape Mixing; negatively for syllogistic reasoning and animacy/size Processing Speed).For the RT-based tasks, the same set of analyses were conducted for accuracies, yielding a qualitatively matching pattern of results (see Table S4 in the Supplemental Material).

In the next step of the analyses, we evaluated whether these occasional effects on single tasks generalize across the multiple indicators for each of the cognitive abilities by running two models. Model 1 includes each bilingualism indicator as a continuous predictor and Model 2 tests the impact of levels of bilingualism as defined by cluster membership. Table 6 lists estimates for Model 1 and Model 2 on each of the cognitive abilities of interest as related to the four hypotheses tested in this study. Complete LME results are listed in Tables S2 and S3

(Supplemental Materials). In both models, the intercept reflecting the baseline (in our case Processing Speed) was non-significant (*p*s > .211). Therefore, we can be confident that effects of bilingualism on single cognitive abilities are not distorted by differences in Processing Speed. As for the correlations, negatively signed estimates reflect a positive impact of bilingualism in the case of age of L2 acquisition, while the opposite is the case for usage and proficiency. The interaction terms with Material were included to test for the possibility that effects may depend on task material (verbal vs. figural-spatial stimuli).

In summary, we found neither support for the *Inhibitory Control Advantage* nor the *Conflict Monitoring Advantage Hypotheses* as none of the language-related predictors had a significant impact on Inhibition, Conflict Speed, Mixing, or WM Monitoring. Regarding the *Shifting Advantage Hypothesis*, we observed a trend for an interaction between shifting, task material, and non-L1 language usage in Model 1 ($b = 0.02$, $p = .068$), and for an interaction between shifting and high level of bilingualism in Model 2 ($b = 0.72$, $p = .078$). Contrary to the hypothesis' predictions, the positive beta loadings suggest that bilingualism was associated with higher switch costs in the figural-spatial (i.e., color-shape) shifting task. Inspecting the task more closely revealed that greater Non-L1 language usage was related to faster RTs, but more strongly so for repetition trials ($r(116) = -.21$, $p = .024$) than for switch trials ($r(116) = -.15$, $p = .110$). Finally, we also found no support for the *Generalized Cognitive Advantage Hypothesis* as neither the continuous indicators of bilingualism nor the cluster membership had a significant main effect on overall cognitive performance. In contrast, the covariate for parents' education was significant in both Model 1 ($b = 0.07$, $p < .001$) and Model 2 ($b = 0.06$, $p < .001$), with higher levels of parents' education predicting better cognitive ability.

Discussion

We found no evidence for any of four hypotheses predicting bilingual benefits, converging with other recent larger-scale investigations failing to observe bilingual advantages (e.g., Antón, et al., 2014; Duñabeitia, et al., 2014; Gathercole, et al., 2014; Hernández, et al., 2013; Paap, et al., 2014). In contrast, our results confirmed the robust impact of parents' education (which served as a proxy for SES) on cognitive performance (for an overview, see Lawson, Hook, Hackman, & Farah, in press). Importantly, our study design accounts for factors discussed to be critical for the detection of bilingual effects (cf. Kroll & Bialystok, 2013). First, we treated bilingualism as a multidimensional, continuous construct. Second, we examined the bilingual advantage hypotheses in multiple cognitive abilities simultaneously in the same, relatively large sample. Finally, the use of linear mixed-effects modeling[3] allowed for controlling multiple sources of variance such as variability in sampling of individuals and tasks, and parents' education.

Limitations

One potential limitation of our study is the lack of a strictly monolingual control group. As our study was conducted in Switzerland, a multilingual country with obligatory second language classes at school, even participants in the language cluster with the lowest level of bilingualism reported at least some exposure to other languages. Therefore, we cannot exclude the possibility that effects of bilingualism may emerge when comparing extreme-groups (i.e., monolinguals without any L2 exposure vs. high-level bilinguals). Treating bilingualism as a continuous phenomenon, however, resulted in considerable variance in the language parameters, and, hence, one should still expect to observe at least some effect of bilingualism. Furthermore, this continuous approach has been strongly advocated as being superior to the use of dichotomous groups (Kroll & Bialystok, 2013). Comparing extreme-groups, in contrast, would yield potential confounds such as that individuals without any L2 exposure will most likely differ from their highly bilingual counterparts in more dimensions than just language. Indeed, differences observed for extreme-groups are likely to vanish once the full sample is analyzed (see Unsworth et al., in press for a recent example).

A second potential limitation of our study is the variety of second languages in our sample. Therefore, we cannot exclude the possibility that bilingual advantages may occur for specific language combinations (e.g., as in Prior & Gollan, 2011). However, none of the four hypotheses of bilingual advantages tested here predicts any such language-specific effects. Instead, advantages are assumed to generalize to any combinations of languages, reflected by the majority of previous studies comprising similarly heterogeneous samples (e.g., Calvo & Bialystok, 2014; Prior & MacWhinney, 2010; Schroeder & Marian, 2012). Furthermore, more recent studies reported no consistent evidence for a major role of language combination for bilingual benefits (e.g., Coderre & van Heuven, 2014; Kaushanskaya & Marian, 2009; Kirk, et al., 2014).

Lastly, we included three highly used indicators of bilingualism in our study which, however, does not exhaust the number of possible indicators that could mediate a relationship between bilingualism and cognitive abilities. For example, we did not assess the frequency of switching between languages as an additional possible indicator of bilingualism (e.g., Prior & Gollan, 2011; Rodriguez-Fornells, Krämer, Lorenzo-Seva, Festman, & Münte, 2012; Soveri, et al., 2011). It is worth noting though that as the number of such mediators increase, the effects of bilingualism on cognitive abilities becomes constrained to highly selective sub-samples of bilinguals.

Implications

We suggest four tentative explanations for the discrepancies between our study together with other recent larger-scale replication failures and earlier studies favoring bilingual advantages: task-specific and sample-specific characteristics, small sample sizes contributing to false-positive findings, and a strong publication bias in the field.

Does the bilingual advantage emerge only for specific tasks and materials? The predominant use of only single indicators to measure cognitive abilities renders task-specific effects to be a likely explanation for inconsistencies between studies (for similar lines of argument, see Paap, et al., 2014; Shipstead, et al., 2012). Therefore, we assessed each construct of interest with multiple indicators that were commonly used in past research and accounted for effects of task materials in our analyses. In fact, if only single tasks were evaluated, occasional task-specific significant correlations between single indicators of bilingualism and cognitive performance were observed (cf. Tables 4 and 5), which could be mistaken for broad effects. Instead, examining variation simultaneously across tasks revealed that the only weak tendency towards any relation between bilingualism and cognitive performance that occurred was highly specific, limited to figural-spatial material (i.e., color-shape shifting), and mainly driven by the proportion of non-L1 language usage. Critically, the color-shape variant is precisely the task that has been used in almost all studies testing the *Shifting Advantage Hypothesis* (e.g., Garbin et al., 2010; Hernández, et al., 2013; Paap & Greenberg, 2013; Paap & Sawi, 2014; Prior & Gollan, 2013; Prior & MacWhinney, 2010). Hence, our observation that this effect does not generalize to other similar tasks suggests that shifting advantages reported previously may have actually been driven by material-specific features.

Are bilingual advantages context-specific? Arguably, bilingual context and awareness vary widely across countries, and are hence factors potentially contributing to diverging findings across labs (cf. Adesope, Lavin, Thompson, & Ungerleider, 2010). For example, bilingual awareness is likely to be higher and social bilingual interactions more frequent in multilingual countries such as Canada or Switzerland compared to more monolingual countries such as Portugal or parts of the United States. Whereas the broader context in which a study is conducted may affect the absence or presence of effects of bilingualism, it is important to keep in mind that the hypotheses tested here do not predict that effects are restricted to specific bilingual contexts or to a certain degree of bilingual awareness. Instead, they predict that a higher degree of bilingualism correlates with better cognitive abilities, which is clearly not supported by our data. Whether and what other factors or circumstances coinciding with bilingualism in some (but not other) contexts may result in the positive effects found in other labs and countries needs to be examined in future studies.

In a similar vein, it has been argued that bilingual advantages may compete with other benefits and may therefore be "invisible" in some samples but not in others (Valian, 2015, p. 3). In our study, we detected no differences in three self-reported leisure activities (musical training, video gaming, or physical exercise) though. Bilingual advantages being masked by other benefits is, therefore, an unlikely explanation for the absence of effects in our study.

Why null-findings matter. Despite employing a methodologically sound design following recent recommendations, we were unable to detect any significant bilingual advantages as they are proposed by the four most prominent hypotheses in the literature (cf. Table 1). We however oppose the view that significant results are "greatly more informative than the attempted replications that fail to find significance" (Kroll & Bialystok, 2013, p. 502). Instead, it is crucial for scientific progress that theory is informed by both positive and null findings, especially so in a field that is at high risk for false-positives publication bias, and vague theorizing. First, as pointed out by Paap (2014), many studies examining bilingual advantages comprised only relatively small samples, with the majority of larger-scale studies failing to detect consistent bilingual advantages (e.g., Antón, et al., 2014; Duñabeitia, et al., 2014; Gathercole, et al., 2014; Hernández, et al., 2013; Paap, et al., 2014; but see Luo, et al., 2013). Small sample sizes are indeed more likely to produce false-positive findings (Button, et al., 2013), the risk of which is additionally increased by research on the bilingual advantage being a particularly 'hot topic' (Ioannidis, 2005) given the desire to report statistically significant results (Simmons, Nelson, & Simonsohn, 2011). Second, systematic biases such as the file-drawer problem (i.e., selectively publishing only significant results) and the confirmation bias (Paap, 2014) potentially aggravate the situation. In fact, de Bruin et al. (2015) presented compelling evidence for a strong publication bias in the field. Third, only few existing theories flesh out the underlying mechanisms of bilingual advantages, while most of the current hypotheses produce rather general predictions (i.e., no effect at all vs. any effect on any measure). Purely verbal and imprecise theories, however, ultimately become non-falsifiable as the absence of evidence can always be explained in terms of an infinite number of possible moderators suppressing or boosting the hypothesized effects. To avoid this issue, theory-driven hypotheses should predict specific situations and constellations for which bilingual advantages are expected, but also such in which no effects are expected to occur (Fiedler, Kutzner, & Krueger, 2012; see also Paap, 2014; Paap & Greenberg, 2013).

Conclusion

Our data revealed no evidence for any bilingual cognitive benefit. Hence, despite the numerous findings favoring bilingual advantages, our study suggests that if existing, cognitive benefits of bilingualism are not as broad and as robust as previously assumed. Considering biases inherent in the field, we argue for the necessity of theories producing falsifiable and more precise predictions, and for more larger-scale investigations of the bilingual advantage.

References

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research, 80*(2), 207-245. doi: 10.3102/0034654310368

Antón, E., Duñabeitia, J. A., Estévez, A., Hernández, J. A., Castillo, A., Fuentes, L. J., . . . Carreiras, M. (2014). Is there a bilingual advantage in the ANT task? Evidence from children. *Frontiers in Psychology*, *5*(398), 1-12. doi: 10.3389/fpsyg.2014.00398

Arthur, W., Jr., & Day, D. V. (1994). Development of a short form for the Raven advanced progressive matrices test. *Educational and Psychological Measurement*, *54*(2), 394-403.

Arthur, W., Jr., Tubre, T. C., Paul, D. S., & Sanchez-Ku, M. L. (1999). College-sample psychometric and normative data on a short form of the Raven advanced progressive matrices test. *Journal of Psychoeducational Assessment*, *17*, 354-361.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 290-412. doi: 10.1016/j.jml.2007.12.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255-278. doi: 10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B. M., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4 (Version 1.1-7).

Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging, 19*(2), 290-303. doi: 10.1037/0882-7974.19.2.290

Bialystok, E., Craik, F. I. M., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(4), 859-873. doi: 10.1037/0278-7393.34.4.859

Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *TRENDS in Cognitive Sciences, 16*(4), 240-250. doi: 10.1016/j.tics.2012.03.001

Bundesamt für Statistik. (2015a). Bevölkerung mit Migrationshintergrund Retrieved 07/29/2015, from http://www.bfs.admin.ch/bfs/portal/de/index/themen/01/07/blank/key/04.html

Bundesamt für Statistik. (2015b). Sprachen Retrieved 07/29/2015, from http://www.bfs.admin.ch/bfs/portal/de/index/themen/01/05/blank/key/sprachen.html

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*, 365-376. doi: 10.1038/nrn3475

Calvo, A., & Bialystok, E. (2014). Independent effects of bilingualism and socioeconomic status on language ability and executive functioning. *Cognition, 130*, 278-288. doi: 10.1016/j.cognition.2013.11.015

Coderre, E. L., & van Heuven, W. J. B. (2014). The effect of script similarity on executive control in bilinguals. *Frontiers in Psychology, 5*(1070), 1-16. doi: 10.3389/fpsyg.2014.01070

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*(5), 739-786. doi: 10.3758/BF03196772

Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition, 113*, 135-145. doi: 10.1016/j.cognition.2009.08.001

Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition, 106*(1), 59-86. doi: 10.1016/j.cognition.2006.12.013

de Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism: An example of publication bias? *Psychological Science, 26*(1), 99-107. doi: 10.1177/0956797614557866

Duñabeitia, J. A., Hernández, J. A., Antón, E., Macizo, P., Estévez, A., Fuentes, L. J., & Carreiras, M. (2014). The inhibitory advantage in bilingual children revisited. *Experimental Psychology, 61*(3), 234-251. doi: 10.1027/1618-3169/a000243

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for Kit of Factor-Referenced Cognitive Tests*. Princeton, NJ: Educational Testing Service.

Engel de Abreu, P. M. J. (2011). Working memory in multilingual children: Is there a bilingual effect? *Memory, 19*(5), 529-537. doi: 10.1080/09658211.2011.590504

Fiedler, K., Kutzner, F., & Krueger, J. I. (2012). The long way from α-error control to validity proper: Problems with a short-sighted false-positive debate. *Psychological Science, 7*(6), 661-669. doi: 10.1177/1745691612462587

Fuller-Thomson, E., & Kuh, D. (2014). The healthy migrant effect may confound the link between bilingualism and delayed onset of Alzheimer's disease. *Cortex, 52*, 128-130. doi: 10.1016/j.cortex.2013.08.009

Garbin, G., Sanjuan, A., Forn, C., Bustamante, J. C., Rodriguez-Pujadas, A., Belloch, V., . . . Ávila, C. (2010). Bridging language and attention: Brain basis of the impact of bilingualism on cognitive control. *NeuroImage, 53*(4), 1272-1278. doi: 10.1016/j.neuroimage.2010.05.078

Gathercole, V. C. M., Thomas, E. M., Kennedy, I., Prys, C., Young, N., Guasch, N. V., . . . Jones, L. (2014). Does language dominance affect cognitive performance in bilinguals? Lifespan evidence from preschoolers through older adults on card sorting, Simon, and metalinguistic tasks. *Frontiers in Psychology, 5*(11), 1-14. doi: 10.3389/fpsyg.2014.00011

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1*(2), 67-81. doi: 10.1017/S1366728998000133

Hakuta, K., & Diaz, R. M. (1985). The relationship between degree of bilingualism and cognitive ability: A critical discussion and some new longitudinal data. In K. E. Nelson (Ed.), *Children's Language* (Vol. 5, pp. 320-344). Hillsdale, NJ: Lawrence Erlbaum.

Hernández, M., Costa, A., Fuentes, L. J., Vivas, L. J., & Sebastián-Gallés, N. (2010). The impact of bilingualism on the executive control and orienting networks of attention. *Bilingualism: Language and Cognition, 13*(3), 315-325. doi: 10.1017/S1366728909990010

Hernández, M., Martin, C. D., Barceló, F., & Costa, A. (2013). Where is the bilingual advantage in task-switching? *Journal of Memory and Language, 69*, 257-276. doi: 10.1016/j.jml.2013.06.004

Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic Bulletin & Review, 18*, 625-658. doi: 10.3758/s13423-011-0116-7

Hilchey, M. D., Saint-Aubin, J., & Klein, R. (in press). Does bilingual exercise enhance cognitive fitness in traditional non-linguistic executive processing tasks? In J. Schwieter (Ed.), *The Cambridge Handbook of Bilingual Processing* (pp. 586-613). Cambridge: Cambridge University Press.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLOS Medicine, 2*(8), e124. doi: 10.1371/journal.pmed.0020124

Ishihara, S. (2003). *Ishihara's tests for colour-blindness*. Tokyo: Kanehara and Co.

Kaushanskaya, M., & Marian, V. (2009). The bilingual advantage in novel word learning. *Psychonomic Bulletin & Review, 16*(4), 705-710. doi: 10.3758/PBR.16.4.705

Kirk, N. W., Fiala, L., Scott-Brown, K. C., & Kempe, V. (2014). No evidence for reduced Simon cost in elderly bilinguals and bidialectals. *Journal of Cognitive Psychology, 26*(6), 640-648. doi: 10.1080/20445911.2014.929580

Kousaie, S., & Phillips, N. A. (2012). Ageing and bilingualism: Absence of a "bilingual advantage" in Stroop interference in a nonimmigrant sample. *The Quarterly Journal of Experimental Psychology*, *65*(2), 356-369. doi: 10.1080/17470218.2011.604788

Kousaie, S., Sheppard, C., Lemieux, M., Monetta, L., & Taler, V. (2014). Executive function and bilingualism in young and older adults. *Frontiers in Behavioral Neuroscience, 8*(250), 1-12. doi: 10.3389/fnbeh.2014.00250

Kroll, J. F., & Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology*, *25*(5), 497-514. doi: 10.1080/20445911.2013.799170

Kroll, J. F., Bobb, S. C., & Hoshino, N. (2014). Two languages in mind: Bilingualism as a tool to investigate language, cognition, and the brain. *Current Directions in Psychological Science, 23*(3), 159-163. doi: 10.1177/0963721414528511

Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2014). lmerTest: Tests for random and fixed effects for linear mixed effects models (lmer objects of lme4 package). Retrieved from http://CRAN.R-project.org/package=lmerTest

Lawson, G. M., Hook, C. J., Hackman, D. A., & Farah, M. J. (in press). Socioeconomic status and neurocognitive development: Executive function. In J. A. Griffin, L. S. Freund & P. McCardle (Eds.), *Executive Function in Preschool Children: Integrating Measurement, Neurodevelopment, and Translational Research*. Washington D.C., MD: American Psychological Association Press.

Lewandowsky, S., Oberauer, K., Yang, L.-X., & Ecker, U. K. H. (2010). A working memory test battery for MATLAB. *Behavior Research Methods, 42*(2), 571-585. doi: 10.3758/BRM.42.2.571

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49*, 764-766. doi: 10.1016/j.jesp.2013.03.013

Logan, G. D., & Zbrodoff, N. J. (1979). When it helps to be misled: Facilitative effects of increasing the frequency of conflicting stimuli in a Stroop-like task. *Memory & Cognition, 7*(3), 166-174. doi: 10.3758/BF03197535

Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology, 25*(5), 605-621. doi: 10.1080/20445911.2013.795574

Luo, L., Craik, F. I. M., Moreno, S., & Bialystok, E. (2013). Bilingualism interacts with domain in a working memory task: Evidence from aging. *Psychology and Aging, 28*(1), 28-34. doi: 10.1037/a0030875

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*(2), 163-203. doi: 10.1037/0033-2909.109.2.163

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*, 940-967.

Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: Language and Cognition, 11*(1), 81-93. doi: 10.1017/S1366728907003227

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*(1), 8-14. doi: 10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49-100. doi: 10.1006/cogp.1999.0734

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*(3), 134-140. doi: 10.1016/S1364-6613(03)00028-7

Morton, J. B., & Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. *Developmental Science, 10*(6), 719-726. doi: 10.1111/j.1467-7687.2007.00623.x

Oberauer, K., Süß, H.-M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity - facets of a cognitive ability construct. *Personality and Individual Differences*, *29*, 1017-1045.

Oberauer, K., Süß, H.-M., Wilhelm, O., & Wittmann, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence, 31*, 167-193. doi: 10.1016/S0160-2896(02)00115-0

Paap, K. R. (2014). The role of componential analysis, categorical hypothesising, replicability and confirmation bias in testing for bilingual advantages in executive functioning. *Journal of Cognitive Psychology, 26*(3), 242-255. doi: 10.1080/20445911.2014.891597

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology, 66*, 232-258. doi: 10.1016/j.cogpsych.2012.12.002

Paap, K. R., Johnson, H. A., & Sawi, O. (2014). Are bilingual advantages dependent upon specific tasks or specific bilingual experiences? *Journal of Cognitive Psychology, 26*(6), 615-639. doi: 10.1080/20445911.2014.944914

Paap, K. R., & Sawi, O. (2014). Bilingual advantages in executive functioning: problems in convergent validity, discriminant validity, and the identification of the theoretical constructs. *Frontiers in Psychology, 5*(962), 1-15. doi: 10.3389/fpsyg.2014.00962

Prior, A., & Gollan, T. H. (2011). Good language-switchers are good task-switchers: Evidence from Spanish-English and Mandarin-English bilinguals. *Journal of the International Neuropsychological Society, 17*, 682-691. doi: 10.1017/S1355617711000580

Prior, A., & Gollan, T. H. (2013). The elusive link between language control and executive control: A case of limited transfer. *Journal of Cognitive Psychology, 25*(5), 622-645. doi: 10.1080/20445911.2013.821993

Prior, A., & MacWhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and Cognition, 13*(2), 253-262. doi: 10.1017/S1366728909990526

R Core Team. (2014). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Raven, J. C. (1990). *Advanced Progressive Matrices: Sets I, II*. Oxford, U.K.: Oxford Psychologists Press.

Rodriguez-Fornells, A., Krämer, U., Lorenzo-Seva, U., Festman, J., & Münte, T. F. (2012). Self-assessment of individual differences in language switching. *Frontiers in Psychology, 2*(388), 1-16. doi: 10.3389/fpsyg.2011.00388

Schroeder, S. R., & Marian, V. (2012). A bilingual advantage for episodic memory in older adults. *Journal of Cognitive Psychology, 24*(5), 591-601. doi: 10.1080/20445911.2012.669367

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin, 138*(4), 628-654. doi: 10.1037/a0027473

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, in press. doi: 10.1177/0956797611417632

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 174-215. doi: 10.1037/0278-7393.6.2.174

Soveri, A., Rodriguez-Fornells, A., & Laine, M. (2011). Is there a relationship between language switching and executive functions in bilingualism? Introducing a within-group analysis approach. *Frontiers in Psychology, 2*(183), 1-8. doi: 10.3389/fpsyg.2011.00183

Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., . . . Bates, E. (2004). A new on-line resource for psycholinguistic studies. *Journal of Memory and Language, 51*(2), 247-250. doi: 10.1016/j.jml.2004.03.002

Unsworth, N., Redick, T. S., McMillan, B. D., Hambrick, D. Z., Kane, M. J., & Engle, R. W. (in press). Is playing videogames related to cognitive abilities? *Psychological Science*.

Valian, V. (2015). Bilingualism and cognition. *Bilingualism: Language and Cognition*, *18*(1), 3-24. doi: 10.1017/S1366728914000522

von Bastian, C. C., Locher, A., & Ruflin, M. (2013). Tatool: A Java-based open-source programming framework for psychological studies. *Behavior Research Methods*, *45*(1), 108-115. doi: 10.3758/s13428-012-0224-y

von Bastian, C. C., & Oberauer, K. (2013). Distinct transfer effects of training different facets of working memory capacity. *Journal of Memory and Language*, *69*, 36-58. doi: 10.1016/j.jml.2013.02.002

Wiseheart, M., Viswanathan, M., & Bialystok, E. (in press). Flexibility in task switching by monolinguals and bilinguals. *Bilingualism: Language and Cognition*. doi: 10.1017/S1366728914000273

Footnotes

[1]In addition to the tasks described below, participants also completed an Antisaccade task which however had to be discarded due to an experimenter error that led to variable viewing distances.

[2]Using congruent trials as the baseline may confound facilitation and interference effects (e.g., see MacLeod, 1991). Therefore, we used neutral trials as the baseline where possible (for a similar reasoning, see Bialystok, et al., 2008; Hernández, Costa, Fuentes, Vivas, & Sebastián-Gallés, 2010). Additional analyses (not reported here) using congruent trials as the baseline condition in the Flanker and Stroop task yielded qualitatively the same patterns of results.

[3]It is possible that continuous effects of bilingualism are not linear, but follow a different relationship. While analyzing the language clusters allows for complex and non-linear relationships between the three indicators of bilingualism, we additionally carefully inspected the scatterplots for all abilities and each indicator, and ran formal tests for quadratic trends. Only two interactions (besides a main effect of parent's education) showed significant quadratic trends: proficiency x monitoring and proficiency x reasoning. These trends have to be interpreted cautiously though. First, the inverted U-shaped function indicated best cognitive performance for a medium level of proficiency, and worse cognitive performance for those individuals with a low or a high level of proficiency, which can hardly be interpreted as a bilingual advantage. Second, these two significant results emerged from a huge number of multiple comparisons (as the model comprised 127 predictors in sum) and would thus not survive appropriate corrections.

Table 1

Four Hypotheses of Bilingual Advantages

Hypothesis	Prediction	Tested Abilities
Inhibitory Control Advantage		
Constantly exerting cross-linguistic inhibitory control results in more efficient inhibitory processes in conflict situations.	Smaller interference effects in inhibition tasks due to faster responses in trials with response conflict.	Inhibition
Conflict Monitoring Advantage		
Constantly monitoring the environment for conflict and the need for resolving this conflict result in more efficient domain-general executive processing in tasks entailing conflict.	Better performance in congruent or neutral situations in tasks with cognitive conflict present.	Separately for Conflict Speed, Mixing, and Working Memory Monitoring
Shifting Advantage		
Constantly switching between languages results in more efficient mental set-shifting.	Smaller switching costs due to faster responses in switch trials.	Shifting
Generalized Cognitive Advantage		
Constantly experiencing bilingualism results in overall better cognitive performance.	Better performance across cognitive tasks independent of task-specific characteristics.	Computed across Inhibition, Conflict Speed, Mixing, Working Memory (WM) Monitoring, Shifting, Updating, WM Capacity, and Reasoning

Table 2

Sample Characteristics

Measure	Full Sample	Language Cluster		
		Low	Medium	High
Sample Size	118	43	47	28
Demographics				
Gender (f/m)	74/44	28/15	29/18	17/11
Age (years)	24.17 (3.62)	23.77 (3.09)	24.09 (3.79)	24.93 (4.06)
Education (years)	15.81 (3.28)	15.23 (2.59)	15.77 (3.47)	16.75 (3.77)
Parents' education (0-8 scale)[a]	5.08 (1.58)	4.80 (1.47)	5.37 (1.70)	5.02 (1.51)
Immigration status (%)	33	16[b]	46	39
Primary language measures				
Age of L2 acquisition (years)	7.64 (4.12)	9.47 (3.09)	7.77 (4.08)	4.18 (3.32)
Non-L1 language usage (%)	24.14 (13.99)	11.07 (3.18)	23.26 (4.35)	45.71 (6.65)
Proficiency ratio (L2/L1)	0.69 (0.22)	0.58 (0.22)	0.70 (0.21)	0.85 (0.17)
Other language measures				
Proficiency in L2 (0-10 scale)	6.72 (2.18)	5.65 (2.09)	6.79 (1.95)	8.25 (1.75)
Learnt L2 at school (%)	67.80	88.37	70.21	32.14
Indicated L2 as native (%)	30.51	6.98	31.91	64.28
Leisure activities (h/week)				
Musical training	0.94 (1.93)	0.60 (1.17)	1.13 (2.21)	1.18 (2.34)
Video gaming	1.54 (2.95)	1.33 (3.29)	1.96 (2.83)	1.20 (2.55)
Physical exercise	2.94 (2.93)	3.72 (3.18)	2.54 (2.49)	2.41 (3.04)

Note. Standard deviations are given in parentheses where applicable. L1 = first language; L2 = second language.

[a]Average of mother's and father's educational degree coded as ranging from 0 to 8.

[b]For 5 out of 7 of these participants, their nationality indicated that they immigrated from neighboring countries where the same language is spoken (i.e., Austria and Germany). The remaining 2 participants indicated Swiss nationality and reported to have moved to Switzerland during their first year of life.

Table 3

Descriptive Statistics for Tasks

Task	Material	*M*	*SD*	Min	Max	# of Trials	Reliability
Inhibition							
Simon	f	-0.19	0.09	-0.45	0.02	200	.80[a]
Flanker	v	-0.04	0.06	-0.29	0.10	144	.49[a]
Stroop	n	-0.08	0.06	-0.22	0.05	144	.18[a]
Conflict Speed							
Simon	f	432	54	331	719	200	.99[b]
Flanker	v	521	57	405	707	144	.91[b]
Stroop	n	673	89	505	941	144	.94[b]
Mixing							
Color/Shape	f	-0.52	0.19	-1.03	-0.04	128	.95[a]
Animacy/Size	v	-0.33	0.16	-0.74	0.01	128	.97[a]
Parity/Magnitude	n	-0.31	0.18	-0.88	0.00	128	.96[a]
Working Memory Monitoring							
Squares	f	2.57	0.39	1.43	3.33	16	.64[a]
Rhymes	v	2.70	0.66	0.80	4.02	16	.71[a]
Digits	n	2.80	0.70	0.86	4.36	16	.71[a]
Shifting							
Color/Shape	f	-0.35	0.17	-0.81	0.06	128	.91[a]
Animacy/Size	v	-0.34	0.16	-0.76	0.04	128	.91[a]
Parity/Magnitude	n	-0.36	0.15	-0.89	0.11	128	.90[a]
Updating							
Color keep-track	f	0.62	0.16	0.25	0.94	25	.92[b]
Letter keep-track	v	0.72	0.11	0.18	0.93	25	.89[b]
Digit keep-track	v	0.74	0.18	0.26	1.00	25	.93[b]
Working Memory Capacity							
Spatial short-term memory	f	0.75	0.08	0.53	0.93	15	.74[b]
Brown-Peterson	v	0.74	0.15	0.15	1.00	16	.92[b]
Complex span	n	0.52	0.15	0.22	0.87	16	.90[b]
Reasoning							
Raven's APM	f	0.69	0.20	0.17	1.00	12 (15 min)	.69[b]
Locations	f	0.59	0.18	0.12	0.94	28 (12 min)	.79[b]
Letter sets	v	0.76	0.11	0.37	0.97	30 (14 min)	.73[b]
Nonsense Syllogisms	v	0.62	0.15	0.24	0.93	30 (8 min)	.75[b]
Diagramming Relationships	v	0.75	0.13	0.40	0.97	30 (8 min)	.73[b]
Processing Speed							
Color/Shape	f	494	55	392	706	4 x 64	.99[b]
Animacy/Size	v	593	72	447	893	4 x 64	.99[b]
Parity/Magnitude	v	547	58	430	786	4 x 64	.90[b]

Note. f = figural; v = verbal; n = numerical material. If applicable, time-limits are given in parentheses after the number of trials.

[a]Split-half reliability corrected using Spearman-Brown's prophecy formula.

[b]Cronbach's alpha.

Table 4

Effects of Continuous Predictors of Bilingualism on the Level of Single Tasks

	AoA		Usage		Proficiency	
Task	*r*	*p*	*r*	*p*	*r*	*p*
			Inhibition			
Simon	-.17	.069	.21	**.023**	.05	.565
Flanker	-.20	**.034**	.02	.867	.10	.258
Stroop	.02	.788	.03	.716	.02	.848
			Conflict Speed			
Simon	-.04	.689	-.02	.866	.01	.876
Flanker	-.06	.514	-.06	.543	-.12	.198
Stroop	-.01	.921	.00	.976	.04	.643
			Mixing			
Color/Shape	-.02	.805	.27	**.003**	.14	.122
Animacy/Size	.05	.573	.05	.598	-.10	.274
Parity/Magnitude	-.06	.544	.10	.272	.13	.152
			Working Memory Monitoring			
Squares	.04	.657	.15	.114	.03	.722
Rhymes	.14	.118	-.08	.386	-.02	.797
Digits	-.02	.819	.08	.367	.04	.655
			Shifting			
Color/Shape	-.06	.546	-.11	.249	-.05	.569
Animacy/Size	.01	.947	.02	.845	.06	.551
Parity/Magnitude	-.05	.617	.09	.359	.01	.880
			Updating			
Color Keep-Track	.02	.844	-.04	.673	.02	.832
Letter Keep-Track	.07	.450	.04	.695	-.03	.777
Digit Keep-Track	-.05	.582	.04	.693	-.03	.781
			Working Memory Capacity			
SSTM	-.09	.358	-.01	.895	-.01	.952
Brown-Peterson	.07	.445	-.01	.892	-.04	.665
Complex Span	.03	.708	-.01	.914	-.05	.621
			Reasoning			
Raven's APM	-.10	.274	-.02	.818	-.06	.539
Locations	-.03	.725	.06	.517	-.02	.857
Letter Sets	-.07	.436	-.01	.902	.00	.972
N. Syllogisms	-0.07	.425	-0.12	.202	-0.19	**.043**
D. Relationships	0.01	.882	-0.02	.868	-0.15	.108
			Processing Speed			
Color/Shape	-0.01	.885	0.06	.538	0.03	.754
Animacy/Size	-0.06	.517	-0.01	.899	0.09	.323
Parity/Magnitude	0.00	.969	-0.06	.522	-0.08	.377

Note. AoA = Age of L2 acquisition; SSTM = spatial short-term memory; APM = advanced progressive matrices; N = nonsense; D = diagramming; f = figural; v = verbal; n = numerical. Significant *p*-values are printed in bold.

Table 5

Descriptive Statistics and Effects of Language Cluster on the Level of Single Tasks

	Language Cluster						Effects of Language Cluster				
	Low		Medium		High		Main Effect		Pairwise Comparisons		
Task	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F*	*p*	L vs. M	L vs. H	M vs. H
Inhibition											
Simon	-0.21	0.09	-0.19	0.09	-0.16	0.08	2.53	.084	.333	.079	.312
Flanker	-0.04	0.05	-0.03	0.07	-0.03	0.05	0.42	.661	> .999	> .999	> .999
Stroop	-0.08	0.07	-0.08	0.06	-0.08	0.06	0.20	.820	> .999	> .999	> .999
Conflict Speed											
Simon	425	52	439	62	431	42	0.70	.499	.720	> .999	> .999
Flanker	515	52	531	60	512	58	1.27	.285	.530	.850	.530
Stroop	671	92	669	83	683	97	0.23	.792	> .999	> .999	> .999
Mixing											
Color/Shape	-0.57	0.19	-0.52	0.19	-0.45	0.18	3.97	**.022**	.192	**.017**	.192
Animacy/Size	-0.34	0.17	-0.34	0.15	-0.32	0.14	0.15	.863	> .999	> .999	> .999
Parity/Magnitude	-0.33	0.22	-0.29	0.16	-0.30	0.15	0.52	.598	> .999	> .999	> .999
Working Memory Monitoring											
Squares	2.53	0.42	2.54	0.37	2.68	0.34	1.51	.225	.880	.330	.330
Rhymes	2.71	0.73	2.76	0.57	2.59	0.70	0.62	.538	.890	.890	.810
Digits	2.83	0.71	2.63	0.68	3.02	0.66	2.97	.055	.336	.336	.054
Shifting											
Color/Shape	-0.34	0.17	-0.34	0.17	-0.39	0.18	0.83	.438	.920	.710	.710
Animacy/Size	-0.35	0.16	-0.32	0.17	-0.35	0.14	0.70	.500	.820	.910	.820
Parity/Magnitude	-0.36	0.15	-0.37	0.16	-0.32	0.15	0.52	.598	.830	.470	.470
Updating											
Color Keep-Track	0.63	0.17	0.62	0.15	0.61	0.17	0.08	.923	> .999	> .999	> .999
Letter Keep-Track	0.71	0.12	0.72	0.12	0.72	0.10	0.08	.919	> .999	> .999	> .999

Digit Keep-Track	0.75	0.19	0.72	0.18	0.77	0.14	0.66	.519	.990	.990	.790
					Working Memory Capacity						
SSTM	0.75	0.07	0.74	0.07	0.76	0.09	0.54	.583	> .999	> .999	> .999
Brown-Peterson	0.74	0.13	0.74	0.17	0.73	0.11	0.10	.906	> .999	> .999	> .999
Complex Span	0.52	0.16	0.53	0.14	0.51	0.17	0.11	.900	> .999	> .999	> .999
					Reasoning						
Raven's APM	0.69	0.20	0.70	0.19	0.68	0.23	0.07	.932	> .999	> .999	> .999
Locations	0.60	0.19	0.57	0.19	0.62	0.14	0.62	.542	> .999	> .999	.840
Letter Sets	0.76	0.14	0.76	0.10	0.75	0.08	0.20	.816	> .999	> .999	> .999
N. Syllogisms	0.64	0.15	0.62	0.16	0.58	0.13	1.31	.273	.560	.330	.530
D. Relationships	0.77	0.15	0.73	0.11	0.77	0.12	0.98	.380	.590	.890	.630
					Processing Speed						
Color/Shape	482	50	502	62	496	49	1.55	.218	.260	.590	.650
Animacy/Size	575	54	614	82	584	72	3.90	**.023**	**.025**	.600	.138
Parity/Magnitude	543	54	558	66	535	43	1.64	.199	.430	.550	.270

Note. AoA = Age of L2 acquisition; L = low; M = medium; H = high; SSTM = spatial short-term memory; APM = advanced progressive matrices; N = nonsense; D = diagramming; f = figural; v = verbal; n = numerical. *P*-values for pairwise comparisons are Holm-adjusted; significant *p*-values are printed bold.

Table 6

Parameter Estimates for Fixed Effects in the Linear Mixed-Effects Models Related to the Four Hypotheses of Bilingual Advantages

	Model 1: Degree of Bilingualism									Model 2: Language Clusters					
	Age of acquisition			Usage			Proficiency			Medium[a]			High[a]		
Cognitive Ability	*b*	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*
Inhibitory Control Advantage Hypothesis															
Inhibition	-0.03	0.03	.415	0.01	0.01	.404	-0.29	0.63	.647	-0.16	0.29	.576	0.28	0.33	.388
Inhibition x Material[b]	0.02	0.04	.609	-0.01	0.01	.619	0.44	0.77	.568	-0.09	0.35	.797	-0.16	0.41	.682
Conflict Monitoring Advantage Hypothesis															
Conflict Speed	-0.01	0.03	.701	0.00	0.01	.666	-0.02	0.63	.972	-0.11	0.29	.691	-0.15	0.33	.661
Conflict Speed x Material	0.01	0.04	.787	0.01	0.01	.609	-0.23	0.77	.760	-0.17	0.35	.630	0.20	0.41	.629
Mixing	0.03	0.03	.402	0.02	0.01	.095	0.21	0.63	.735	-0.09	0.29	.764	0.41	0.33	.214
Mixing x Material	-0.01	0.04	.899	-0.01	0.01	.652	-0.34	0.77	.659	-0.21	0.35	.550	-0.25	0.41	.533
Monitoring	0.03	0.03	.413	0.01	0.01	.310	-0.10	0.63	.872	-0.33	0.29	.251	0.13	0.33	.684
Monitoring x Material	0.01	0.04	.894	0.00	0.01	.772	0.13	0.77	.864	-0.18	0.35	.608	-0.08	0.41	.842
Shifting Advantage Hypothesis															
Shifting	-0.04	0.03	.279	-0.02	0.01	.144	-0.18	0.63	.776	-0.38	0.29	.182	-0.55	0.33	.100
Shifting x Material	0.05	0.04	.208	0.02	0.01	.068	0.19	0.77	.807	0.07	0.35	.842	0.72	0.41	.078
Generalized Cognitive Advantage Hypothesis															
General performance	0.01	0.03	.840	0.00	0.01	.524	-0.05	0.46	.915	0.33	0.21	.119	0.24	0.24	.316
Material	-0.02	0.03	.539	-0.01	0.01	.315	0.03	0.54	.961	0.04	0.25	.859	-0.26	0.29	.356

[a] Contrasted against language cluster with a low level of bilingualism.

[b] Tasks with figural-spatial material contrasted against tasks with verbal-numerical material.

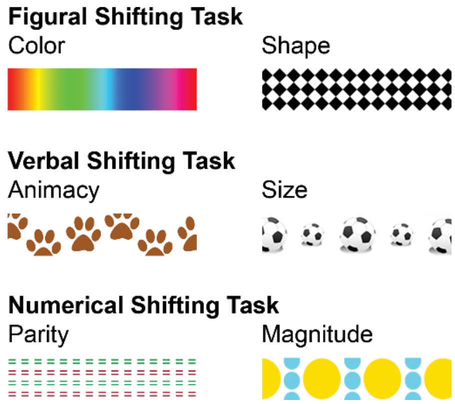

Figure 1. Visual cues indicating the currently relevant rule in the shifting tasks.

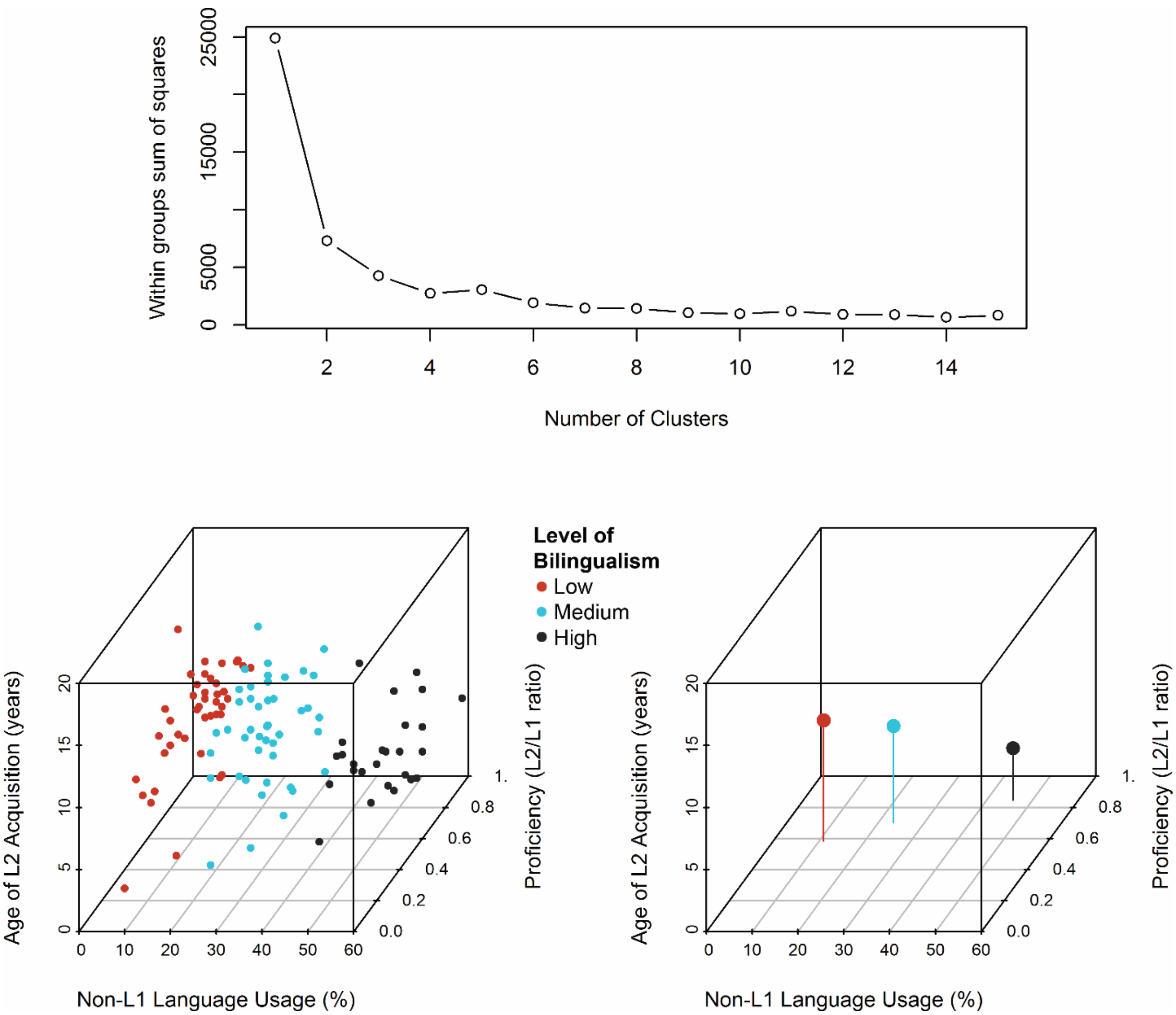

Figure 2. Results of the cluster analysis. Scree plot (top), language cluster scatter plot (bottom left) and cluster means (bottom right).